

Linear Detrending Subsequence Matching in Time-Series Databases

Myeong-Seon Gil, Yang-Sae Moon, and Bum-Soo Kim

Department of Computer Science, Kangwon National University, Korea.
{gils,ysmoon,bskim}@kangwon.ac.kr

Abstract. Each time-series has its own linear trend, the directionality of a time-series, and removing the linear trend is crucial to get the more intuitive matching results. Supporting the linear detrending in subsequence matching is a challenging problem due to a huge number of possible subsequences. In this paper we define this problem the *linear detrending subsequence matching* and propose its efficient *index-based* solution. To this end, we first present a notion of *LD-windows* (LD means linear detrending), which is obtained as follows: we eliminate the linear trend from a subsequence rather than each window itself and obtain LD-windows by dividing the subsequence into windows. Using the LD-windows we then present a lower bounding theorem for the index-based matching solution and formally prove its correctness. Based on the lower bounding theorem, we next propose the index building and subsequence matching algorithms for linear detrending subsequence matching. We finally show the superiority of our index-based solution through extensive experiments.

Key words: data mining, time-series databases, similar sequence matching, linear detrending, subsequence matching

1 Introduction

Time-series data are of growing importance in data mining and data warehousing [7,10]. A time-series is a sequence of real numbers representing values at specific points in time. Typical examples include stock prices, music data, network traffic data, moving object trajectories, and biomedical data [2,3,5,9,16]. The time-series data stored in a database are called *data sequences*, and those given by users are called *query sequences*. Finding data sequences similar to the given query sequence from the database is called *similar sequence matching* or *time-series matching* [7]. In many similar sequence matching models, two sequences $X = \{X[1], \dots, X[n]\}$ and $Y = \{Y[1], \dots, Y[n]\}$ are said to be *similar* if the distance $D(X, Y) \leq \epsilon$, where ϵ is the user-specified tolerance. In this paper we use the Euclidean distance ($= \sqrt{\sum_{i=1}^n |X[i] - Y[i]|^2}$) as the distance function of $D(X, Y)$.

Linear trend, a representative distortion of time-series data [9,12], shows the directionality of a time-series, and *linear detrending* in similar sequence matching is crucial to get the more intuitive matching results. Figure 1 shows an example of comparing two sequences before and after linear detrending: Figure 1(a) represents the original sequences Q and S ; Figure 1(b) the linear detrended sequences

Q' and S' . We obtain Q' and S' by linear detrending, i.e., by subtracting the corresponding trend lines $f(Q)$ and $f(S)$ from Q and S , respectively. In Figure 1, there is a big distance between Q and S , and these two sequences will be determined to be non-similar. In contrast, the distance between Q' and S' is very small in Figure 1(b), and they will be determined to be similar. It means that non-similar sequences can be identified as similar ones after linear detrending, and vice versa. Likewise, linear detrending is useful to know similarity of changes which is hidden by the linear trend of time-series data [8,9]. Motivated by this example, we attack the problem of linear detrending in similar sequence matching, especially in subsequence matching [5,13].

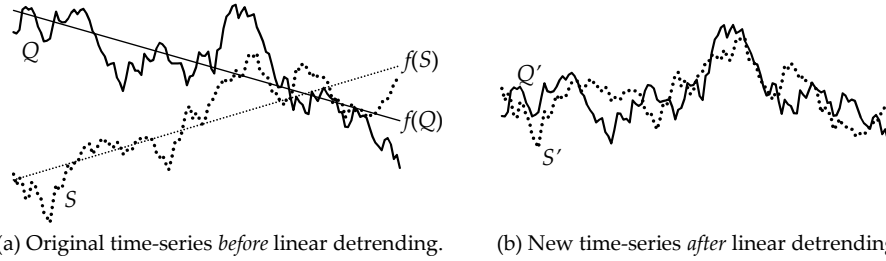


Fig. 1. Comparison of two sequences S and Q *before* and *after* linear detrending.

In this paper we address the problem of linear detrending in subsequence matching. Supporting the linear detrending is simple in whole matching since all data and query sequences have the same length. But, it is a challenging problem in subsequence matching because we need to consider a huge number of possible data subsequences to be linear detrended. We call this matching scheme the *linear detrending subsequence matching*. Formally speaking, for a query sequence Q and a data sequence S , linear detrending subsequence matching finds all subsequences $S[i : j]$ such that $D(\overline{Q}, \overline{S[i : j]}) \leq \epsilon$, where \overline{Q} and $\overline{S[i : j]}$ are the linear detrended (sub)sequences of Q and $S[i : j]$, respectively.

We propose an *index-based* solution for linear detrending subsequence matching. To this end, we first present a novel notion of *LD-windows*, linear detrending-windows. Suppose a subsequence $S[i : j]$ include a window $S[a : b]$ (i.e., $i \leq a < b \leq j$), then we obtain the LD-window of $S[a : b]$ by eliminating the linearity of subsequence $S[i : j]$ rather than that of window $S[a : b]$ itself. This notion enables an LD-window to represent multiple subsequences of different lengths, and eventually, we can use only one index in subsequence matching [15]. Using the LD-windows we next present a lower bounding theorem for the index-based matching solution and formally prove its correctness. Based on this lower bounding theorem, we then propose the index building and subsequence matching algorithms, respectively. We finally showcase the superiority of our index-based solution through extensive experiments. Experimental results show that, com-

pared with the sequential scan, our solution improves the matching performance by one or two orders of magnitude.

2 Related Work

Similar sequence matching can be classified into whole matching and subsequence matching [13]. The whole matching [1,4] finds data sequences similar to a query sequence, where the lengths of data and query sequences are all identical. The subsequence matching [5,7,11,12,14,15] finds subsequences, contained in data sequences, similar to a query sequence of arbitrary length. Likewise, subsequence matching is a generalization of whole matching [5,13], and we thus focus on subsequence matching in this paper.

Many efficient solutions have been proposed for subsequence matching [5,7,13]. These solutions consists of index-building and subsequence matching algorithms. In the index-building algorithm, the solution constructs an R^* -tree as follows: it divides data sequences into multiple windows of size ω ; transforms those windows to $f(\ll \omega)$ -dimensional points using the lower-dimensional transformation such as discrete Fourier transform (DFT) and piecewise aggregate approximation (PAA); and stores the points (or minimum bounding rectangles (MBRs) containing multiple points) into the R^* -tree. In the subsequence matching algorithm, the solution finds similar subsequences as follows: it divides the query sequence into multiple windows of the same size ω ; transforms each window to an f -dimensional point; makes a range query using the point and the tolerance; constructs a candidate set by searching the R^* -tree; and finally obtain actual similar subsequences by eliminating *false alarms* through the post-processing step [1,5,12].

Representative distortions embedded in time-series are offset translation, amplitude scaling, noise, and linear trend [6,9]. In similar sequence matching, there have been many efforts to remove these distortions from time-series data. For example, offset translation and amplitude scaling can be solved by the normalization transform, and its subsequence matching solutions were proposed in [11,15,12]. Also, the moving average transform can alleviate noise of time-series, and its subsequence matching solution was proposed in [14]. To our best knowledge, however, there is no solution to linear detrending subsequence matching, and in this paper we define the problem first and present an efficient index-based solution.

3 Linear Detrending Subsequence Matching

3.1 Problem Definition

For a time-series, its linear trend is a straight line that most likely reflects its directionality. The *least square method* is most widely used to obtain the line of a time-series [8]. For a sequence $X = \{X[1], \dots, X[n]\}$, a linear function by the

least square method is given by $g(k) = \alpha k + \beta$, where α and β are obtained by Eq. (1) [8].

$$\begin{aligned}\alpha &= \frac{n \sum_{k=1}^n kX[k] - \sum_{k=1}^n k \cdot \sum_{k=1}^n X[k]}{n \sum_{k=1}^n k^2 - (\sum_{k=1}^n k)^2}, \\ \beta &= \frac{\sum_{k=1}^n X[k]}{n} - \alpha \frac{\sum_{k=1}^n k}{n}.\end{aligned}\tag{1}$$

Linear detrending is the process of obtaining a new time-series from an original time-series by removing the corresponding linear trend. The following is the formal definition of linear detrending.

Definition 1. For a sequence $X = \{X[1], \dots, X[n]\}$ and its trend line $g(k) = \alpha k + \beta$, the *linear detrending sequence* of X , *LD-sequence* of X in short, is defined as $\overline{X} = \{\overline{X}[1], \dots, \overline{X}[n]\}$, where $\overline{X}[k] = X[k] - g(k)$, $k = 1, 2, \dots, n$. \square

Linear detrending is simply solved in whole matching, but it is a challenging problem in subsequence matching. In whole matching, the lengths of data and query sequences are all identical, and we simply use the whole matching solution [1] after linear detrending of all time-series. In contrast, the solution is not simple in subsequence matching by the following reasons: (1) data subsequences in different positions have different linear trend even though they have the same length; and (2) data subsequences of different lengths also have different linear trend even though they start from the same position. Therefore, we need to consider different linear trend for all possible query lengths and for all possible positions, and we cannot use the traditional whole/subsequence matching solutions for linear detrending subsequence matching.

We formally define the problem of linear detrending subsequence matching. We first present similarity of two sequences by considering the linear detrending.

Definition 2. For two sequences X and Y of the same length and their LD-sequences \overline{X} and \overline{Y} , we define that X and Y (or \overline{X} and \overline{Y}) are *LD-similar* if the Euclidean distance between \overline{X} and \overline{Y} is less than or equal to the tolerance ϵ , i.e., if $D(\overline{X}, \overline{Y}) \leq \epsilon$. \square

Using the concept of LD-similarity, we now define the problem of linear detrending subsequence matching as follows:

Definition 3. For a data sequence S , a query sequence Q , and the tolerance ϵ , *linear detrending subsequence matching* is the problem of finding all subsequences $S[i : j]$ which are LD-similar to Q , i.e., finding all subsequences $S[i : j]$ such that $D(\overline{Q}, \overline{S[i : j]}) \leq \epsilon$. \square

3.2 Sequential Scan-based Solution and Its Problems

Sequential scan accesses every subsequence $S[i : j]$ sequentially and investigates its LD-similarity by computing $D(\overline{Q}, \overline{S[i : j]})$. Algorithm 1 shows the sequential

Algorithm 1 LDSeqScan(data sequence S , query sequence Q , tolerance ϵ)

- 1: Compute a trend line $g(k)$ from Q using the least square method;
 - 2: Obtain \overline{Q} from Q and $g(k)$ through linear detrending;
 - 3: **for each** subsequence $S[i : j]$ of length $Len(Q)$ **do**;
 - 4: Compute a trend line $g'(k)$ from $S[i : j]$ using the least square method;
 - 5: Obtain $\overline{S[i : j]}$ from $S[i : j]$ and $g'(k)$ through linear detrending;
 - 6: Return the subsequence $S[i : j]$ if $D(\overline{Q}, \overline{S[i : j]}) \leq \epsilon$; // LD-similar
 - 7: **end-for**
-

scan algorithm, LDSeqScan, which is simple and self-explained. Algorithm LDSeqScan accesses all possible subsequences one by one in Lines 3 to 7 and returns LD-similar subsequences by investigating the LD-similarity.

The sequential scan algorithm has an advantage of simplicity, but has a disadvantage of incurring severe CPU and I/O overhead. First, the algorithm causes many disk accesses since it accesses an entire data sequence in a database. Second, the algorithm also causes severe CPU overhead since it investigates the LD-similarity for every individual subsequence by performing the linear detrending and by computing the Euclidean distance. This CPU and I/O overhead makes LDSeqScan impractical for a large time-series database. To solve this problem, we propose an efficient index-based solution in the next Section 3.3.

3.3 Index-based Solution and Its Algorithms

As in the traditional subsequence matching [5,7,13], we use the window construction mechanism that divides data and query sequences into disjoint/sliding windows of the fixed size. However, our solution quite differs from the traditional ones in constructing windows due to use of linear detrending. Each window should be mapped to multiple windows in the linear detrending subsequence matching while it does not in the traditional subsequence matching. This is because, in linear detrending subsequence matching, each window has multiple trend lines by different lengths and different positions of subsequences that include the window. Formally speaking, for a given window $S[a : b]$, there are many different subsequences $S[i : j]$'s that include $S[a : b]$; their trend lines are also different from each other; and the window $S[a : b]$ is mapped to multiple windows due to different trend lines. We call this complex property the *multiple mapping property*, which was already presented in the normalization-transformed subsequence matching [15]. The traditional subsequence matching solutions [5,7,13] do not have the multiple mapping property, but we need to support this property in linear detrending subsequence matching.

To support the multiple mapping property, for a given window, we do not remove the linear trend of the window itself, but we instead remove the linear trend of a subsequence including that window. To this end, we present a notion of LD-windows as follows:

Definition 4. Suppose $S[i : j]$ be a subsequence of a sequence S , $g(k)$ be a linear function of $S[i : j]$, and $S[a : b]$ be a window included in $S[i : j]$, then

LD-windows of $S[a : b]$ against $S[i : j]$, denoted by $\overline{S_{\{i,j\}}[a : b]}$, is defined as a new window whose entry $\overline{S_{\{i,j\}}[k]} (k = a, a + 1, \dots, b)$ is set to $S[k] - g(k)$. \square

Definition 4 means that a window $S[a : b]$ is mapped to an LD-window $\overline{S_{\{i,j\}}[a : b]}$ by the trend line of a subsequence $S[i : j]$ which includes $S[a : b]$. Because of the multiple mapping property, there are many subsequences $S[i : j]$'s that include $S[a : b]$, and thus, each window $S[a : b]$ is mapped to multiple LD-windows $\overline{S_{\{i,j\}}[a : b]}$'s for different subsequences $S[i : j]$'s.

Like the traditional subsequence matching algorithms [5, 7, 13], our index-based solution first transforms each *high-dimensional* window to a *low-dimensional* point and then stores the point into the multidimensional index. Unlike the traditional algorithms, however, our solution maps each high-dimensional window to a low-dimensional MBR that bounds multiple low-dimensional points. This is due to the multiple mapping property that a window is mapped to multiple LD-windows. Constructing an MBR from a window is performed as follows: (1) the given window is mapped to multiple LD-windows; (2) LD-windows are transformed to low-dimensional points by the lower-dimensional transformation; and (3) a low-dimensional MBR is constructed by bounding the transformed points. We call this MBR *LD-MBR* and formally define it as follows:

Definition 5. Suppose s be a window of a sequence S , \mathbb{S} be $\{\overline{s} \mid \overline{s} \text{ is an LD window of } s\}$, and $T(\cdot)$ be a function of lower-dimensional transformation, then *LD-MBR* of s , denoted by $\mathbb{M}(T(\mathbb{S}))$, is defined as a low-dimensional MBR that bounds all low-dimensional points $T(\overline{s})$ for all $\overline{s} \in \mathbb{S}$. \square

Figure 2 shows the process of constructing an LD-MBR of a window $S[a : b]$. The process is as follows: (1) the window $S[a : b]$ is mapped to multiple LD-windows $\overline{S_{\{i_k:j_k\}}[a : b]}$'s by considering possible subsequences $S[i_k : j_k]$'s; (2) each LD-window is transformed to a low-dimensional point; and (3) an LD-MBR is constructed by bounding those points.

Our index-based solution is developed from the following Theorem 1.

Theorem 1. For a query sequence Q , a data subsequence $S[i : j]$, a tolerance ϵ , a function $T(\cdot)$ of lower-dimensional transformation, if Q and $S[i : j]$ are LD-similar, that is, if $D(\overline{Q}, \overline{S[i : j]}) \leq \epsilon$, the distance between $T(\overline{q_k})$ and $\mathbb{M}(T(\mathbb{S}_k)) \leq \epsilon/\sqrt{p}$, where $\overline{s_1}, \dots, \overline{s_p}$ and $\overline{q_1}, \dots, \overline{q_p}$ are p disjoint windows of \overline{Q} and $\overline{S[i : j]}$, respectively, and \mathbb{S}_k is the set of LD-windows of s_k . That is, the following Eq. (2) holds:

$$D(\overline{Q}, \overline{S[i : j]}) \leq \epsilon \implies \bigvee_{k=1}^p D(T(\overline{q_k}), \mathbb{M}(T(\mathbb{S}_k))) \leq \epsilon/\sqrt{p}. \quad (2)$$

PROOF: The proof is similar to that of *normalization-transformed subsequence matching* in the previous work [15]. Refer to the proof of Theorem 1 in [15] for the detailed proof. \square

Theorem 1 guarantees correctness of our index-based solution to linear detrending subsequence matching. Like the traditional subsequence matching solutions,

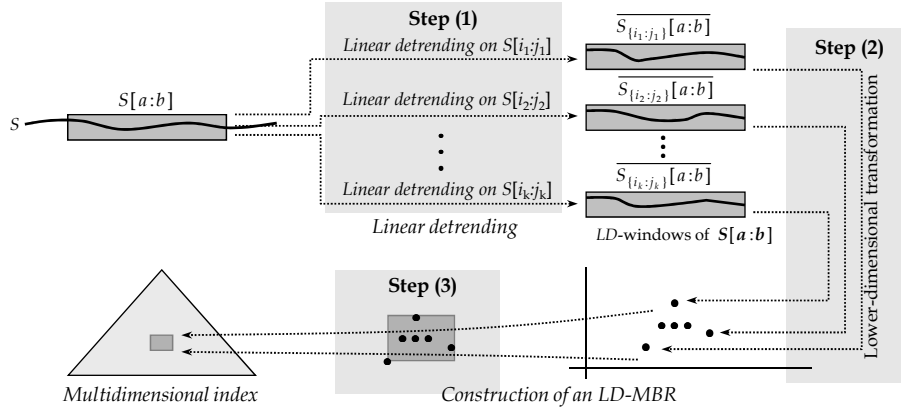


Fig. 2. Process of constructing an LD-MBR of a window $S[a : b]$.

our solution also consists of two algorithms: (1) the index-building algorithm and (2) the subsequence matching algorithm.

Algorithm 2 shows the index-building algorithm. In Line 2 we divide the given data sequence into sliding or disjoint windows of size ω . For the first subsequence matching solution of [5], we use the sliding window; in contrast, for the recent Dual Match [13], we use the disjoint window. In Lines 4 to 14, we build a multidimensional index by repeating the following three steps for each window $S[a : b]$: (1) compute trend lines of all possible subsequences (Line 8); obtain LD-windows using those trend lines (Line 9); and (3) map those LD-windows to an LD-MBR (Line 10). After obtaining an LD-MBR from a window, we store it into the index with its starting offset a (Line 13). Once we build an index by Algorithm BuildIndex, we use it repeatedly in the subsequence matching algorithm.

Algorithm 3 shows the subsequence matching algorithm. In Line 2 we first eliminate the linear trend from the query sequence Q . In Line 3 we divide the LD sequence \bar{Q} into disjoint or sliding windows \bar{q} of size ω . For the first solution of [5], we use the disjoint window; in contrast, for Dual Match [13], we use the sliding window. In Lines 5 to 11, we find candidate subsequences by repeating the following steps for each query window \bar{q} : (1) transform a high-dimensional window \bar{q} to a low-dimensional point (Line 6); (2) make a range query using that point and the given tolerance (Line 7); and (3) find candidate subsequences by evaluating the range query on the index (Lines 8 and 9). After obtaining the candidate subsequences, we finally perform the post-processing step [1,5,7,13] to identify true LD-similar subsequences by accessing actual subsequences and eliminating false alarms.

Algorithm 2 BuildIndex(data sequence S)

```
1: Let the window size be  $\omega$  and the maximum/minimum query lengths be  $l_{min}, l_{max}$ ;
2: Divide  $S$  into windows of size  $\omega$ ;
3: // sliding windows for [5]; disjoint windows for Dual Match [13].
4: for each window  $S[a : b]$  in  $S$  do
5:   Make an  $f$ -dimensional MBR  $\mathbb{M}$  which is initially empty;
6:   for each query length  $l \in [l_{min}, l_{max}]$  do
7:     for each subsequence  $S[i : j]$  of length  $l$  that includes  $S[a : b]$  do
8:       Compute a trend line of  $S[i : j]$  based on the least square method;
9:       Obtain the LD-window  $\overline{S_{\{i,j\}}[a : b]}$ ; // linear detrending
10:      Transform  $\overline{S_{\{i,j\}}[a : b]}$  to an  $f$ -dimensional point and include it into  $\mathbb{M}$ ;
11:     end-for
12:   end-for
13:   Make a record  $\langle \mathbb{M}, \text{offset} = a \rangle$  for  $S[a : b]$ , and store it into the index;
14: end-for
```

Algorithm 3 SubsequenceMatching(query sequence Q , tolerance ω)

```
1: Let the window size be  $\omega$ ; //  $\omega$  is the same one used in Algorithm 2.
2: Obtain  $\overline{Q}$  from  $Q$  by eliminating the linear trend;
3: Divide  $\overline{Q}$  into windows of size  $\omega$ ;
4: // disjoint windows for [5]; sliding windows for Dual Match [13].
5: for each window  $\overline{q}$  do
6:   Transform  $\overline{q}$  to an  $f$ -dimensional point; // lower-dimensional transformation
7:   Construct a range query using that point and  $\epsilon/\sqrt{p}$ ;
8:   //  $p$  is the number of included windows in  $Q$  [13].
9:   Evaluate the query on the index and find the record of the form  $\langle \mathbb{M}, a \rangle$ ;
10:  Include in the candidate set the subsequences  $S[i : j]$  obtained from  $\langle \mathbb{M}, a \rangle$ ;
11: end-for
12: Perform the post-processing step [1,5,7,13] to eliminate false alarms;
```

4 Experimental Evaluation

4.1 Experimental Setup

We have performed experiments using three real data sets, which also used in the previous work [9]. A data set consists of a long data sequence and has the same effect as the one consisting of multiple data sequences [5,13]. The first data set contains electrocardiogram (ECG) data, and we call this data set *ECG-DATA*. The second data set shows tax growth rates, and we call this data set *TAX-DATA*. The third data set contains exchange rate data, and we call this data set *EXCH-DATA*. The length of each data set is 100,000, that is, each data set consists of 100,000 entries (time points).

In the experiment we have compared two matching solutions: (1) LDSeqScan, a sequential scan solution presented in Section 3.2; (2) an index-based matching solution proposed in Section 3.3. We have adopted the first subsequence matching solution of [5] and implemented our index-based approach to that

subsequence matching solution. For simplicity, we call this index-based solution **LDIndexMatch**. We have performed two different experiments. In the first experiment we set the window size and the selectivity¹ [5,13] to 256 and 10^{-3} , respectively, and vary the query sequence length from 256 to 1024. In the second experiment we set the window size and the query sequence length to 256 and 512, respectively, and vary the selectivity from 10^{-1} to 10^{-4} . We obtain the desired selectivity by controlling the tolerance ϵ [13]. As the metric of efficiency, we measure the elapsed time of each solution. We generate query sequences from the data sequence by taking subsequences of length $Len(Q)$ starting from random offsets [5,13,15]. To avoid effects of noise, we experiments with 20 different query sequences of the same length and use their average as the result.

The hardware platform was SUN Ultra 25 workstation equipped with Ultra-SPARC IIIi 1.34GHz CPU, 1.0GB RAM, and an 80GB hard disk; its software platform was Solaris 10. We used C/C++ language for implementing two matching solutions. In **LDIndexMatch**, we used PAA [7,9] as the lower-dimensional transformation and extracted eight features from an window of size 256. As the multidimensional index, we used the R*-tree [1,5,13] for **LDIndexMatch**.

4.2 Experimental Results

Figure 3 shows the results of the first experiment that uses different lengths of query sequences. We first note that, in Figure 3(a) of ECG-DATA, **LDIndexMatch** significantly outperforms **LDSeqScan**. This means that the notion of LD-windows works properly, and it prunes many unnecessary accesses on subsequences at the index level. As shown in Figure 3(a), as the query sequence length decreases, the performance difference between two solutions becomes larger. For example, compared with **LDSeqScan**, **LDIndexMatch** reduces the elapsed time by 38.0 times for the query sequence of length 1024; in contrast, it reduces the elapsed time by 1.60 times only for the query sequence of length 256. This is explained by the *window size effect* [13] that the performance of index-based solutions decreases as the query sequence length on the given window size increases. We can solve this problem by using the *index interpolation* technique [11] which uses multiple indexes (for multiple window sizes) to obtain the better performance. Figures 3(b) and 3(c) of TAX-DATA and EXCH-DATA show the very similar trend with Figure 3(a) of ECG-DATA. It means that the proposed **LDIndexMatch** exploits the pruning effect efficiently, regardless of data types. In summary of Figure 3, our index-based solution, **LDIndexMatch**, improves the overall performance by 1.57 to 38.0 times compared with the straightforward solution, **LDSeqScan**.

Figure 4 shows the results of the second experiment that uses difference selectivities (i.e., different tolerances). As in Figure 3, **LDIndexMatch** also outperforms **LDSeqScan** in all selectivity ranges of Figure 4. We note that, in Figure 4(a) of ECG-DATA, the performance difference between **LDIndexMatch** and **LDSeqScan** increases as the selectivity decreases. This is because, as the selectivity decreases, the number of candidate subsequences also decreases. More precisely, as shown

¹ Selectivity = $\frac{\text{the number of subsequences that are LD-similar with the query sequence}}{\text{the number of all possible subsequences in the database}}$

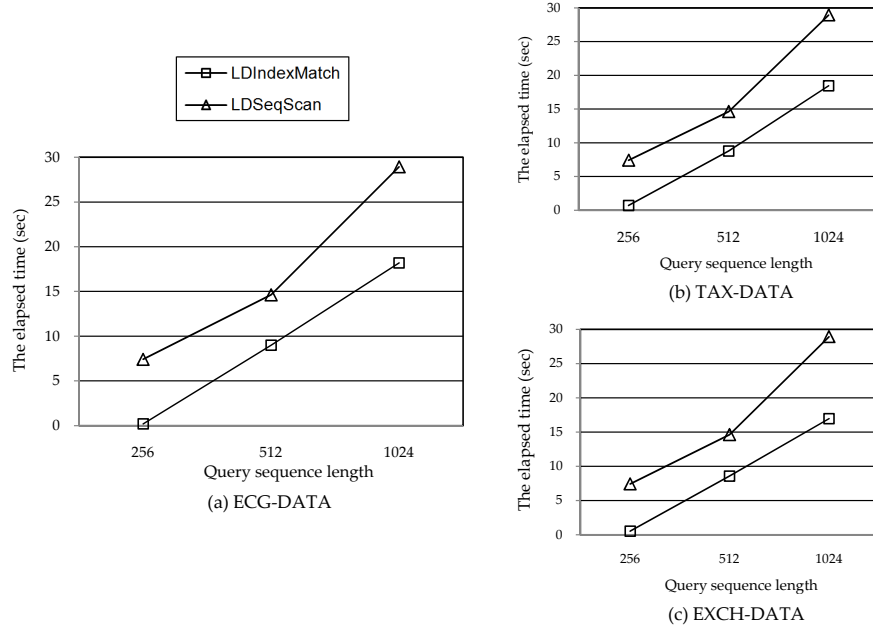


Fig. 3. Experimental results by varying the query sequence length.

in Lines 7 to 10 of Algorithm 3, the smaller selectivity incurs the smaller number of candidate subsequences at the index level, and this reduces the false alarms that cause the disk accesses and the expensive computations on the actual subsequences. As in Figure 3, Figures 4(b) and 4(c) of TAX-DATA and EXCH-DATA show the very similar trend with Figure 4(a) of ECG-DATA. Figure 4 demonstrates that LDIndexMatch beats LDSeqScan in all selectivity ranges, and this means that LDIndexMatch does not much depend on the selectivity values.

5 Conclusions

In this paper we introduced a new problem of linear detrending subsequence matching and proposed an efficient index-based solution. Contributions of the paper are summarized as follows. First, we formally defined the linear detrending subsequence matching and presented its sequential scan-based solution. Second, we presented a novel notion of *LD-windows*, and using LD-windows we proposed an index-based solution. We here formally proved correctness of our index-based solution. Third, we described the index-building and subsequence matching algorithms of the index-based solution. Fourth, we showcased that, compared with the straightforward sequential scan, our index-based solution significantly improved the matching performance by one or two orders of magnitude. We believe that the linear detrending subsequence matching and its index-based solution

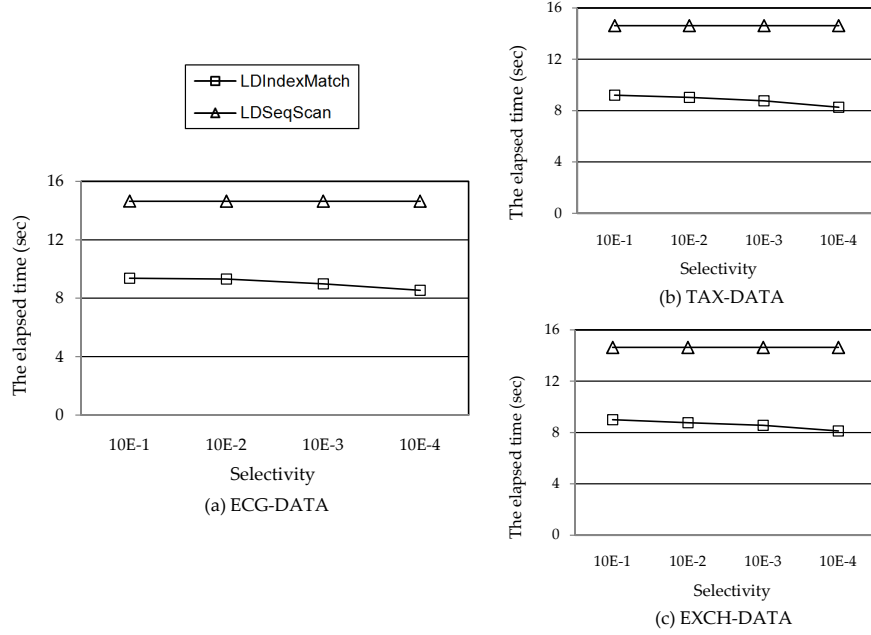


Fig. 4. Experimental results by varying the selectivity.

will be very helpful to find meaningful time-series patterns hidden by the linear trend.

Acknowledgments

This work was partially supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract. (UD060048AD)

References

1. Agrawal, R., Faloutsos, C., and Swami, A., "Efficient Similarity Search in Sequence Databases," In *Proc. of the 4th Int'l Conf. on Foundations of Data Organization and Algorithms*, Chicago, Illinois, pp. 69-84, Oct. 1993.
2. Barrena, M., Jurado, E., Marquez-Neila, P., and Pachon, C., "A Flexible Framework to Ease Nearest Neighbor Search in Multidimensional Data Spaces," *Data and Knowledge Engineering*, Vol. 69, No. 1, pp. 116-136, Jan. 2010.
3. Bu, Y., Chen, L., Fu, A. W.-C., and Liu, D., "Efficient Anomaly Monitoring over Moving Object Trajectory Streams," In *Proc. of the 15th Int'l Conf. on Knowledge Discovery and Data Mining*, ACM SIGKDD, Paris, France, pp. 159-168, June 2009.
4. Chan, K.-P., Fu, A. W.-C., and Yu, C. T., "Haar Wavelets for Efficient Similarity Search of Time-Series: With and Without Time Warping," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 15, No. 3, pp. 686-705, Jan./Feb. 2003.

5. Faloutsos, C., Ranganathan, M., and Manolopoulos, Y., "Fast Subsequence Matching in Time-Series Databases," In *Proc. of Int'l Conf. on Management of Data*, ACM SIGMOD, Minneapolis, Minnesota, pp. 419-429, May 1994.
6. Fu, A. W.-C., Keogh, E. J., Lau, L. Y. H., Ratanamahatana, C., and Wong, R. C.-W., "Scaling and Time Warping in Time Series Querying," *The VLDB Journal*, Vol. 17, No. 4, pp. 899-921, July 2008.
7. Han, W.-S., Lee, J., Moon, Y.-S., and Jiang, H., "Ranked Subsequence Matching in Time-Series Databases," In *Proc. of the 33rd Int'l Conf. on Very Large Data Bases*, Vienna, Austria, pp. 423-434, Sept. 2007.
8. Hill, R. C., Griffiths, W. E., and Judge, G. G., *Undergraduate Econometrics*, John Wiley & Sons Inc., 2007.
9. Keogh, E., "A Decade of Progress in Indexing and Mining Large Time Series Databases," In *Proc. of the 32nd Int'l Conf. on Very Large Data Bases*, Seoul, Korea, p. 1268, Sept. 2006.
10. Lian, X., Chen, L., Yu, J. X., Han, J., and Ma, J. "Multiscale Representations for Fast Pattern Matching in Stream Time Series," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 21, No. 4, pp. 568-581, Apr. 2009.
11. Loh, W.-K., Kim, S.-W., and Whang, K.-Y., "A Subsequence Matching Algorithm that Supports Normalization Transform in Time-Series Databases," *Data Mining and Knowledge Discovery*, Vol. 9, No. 1, pp. 5-28, July 2004.
12. Loh, W.-K., Moon, Y.-S., and Srivastava, J., "Distortion-free Predictive Streaming Time-Series Matching," *Information Sciences*, Vol. 180, No. 8, pp. 1458-1476, Apr. 2010.
13. Moon, Y.-S., Whang, K.-Y., and Han, W.-S., "General Match: A Subsequence Matching Method in Time-Series Databases Based on Generalized Windows," In *Proc. Int'l Conf. on Management of Data*, ACM SIGMOD, Madison, Wisconsin, pp. 382-393, June 2002.
14. Moon, Y.-S. and Kim, J., "Efficient Moving Average Transform-Based Subsequence Matching Algorithms in Time-Series Databases," *Information Sciences*, Vol. 177, No. 23, pp. 5415-5431, Dec. 2007.
15. Moon, Y.-S. and Kim, J., "Fast Normalization-Transformed Subsequence Matching in Time-Series Databases," *IEICE Trans. on Information and Systems*, Vol. E90-D, No. 12, pp. 2007-2018, Dec. 2007.
16. Yi, B.-K., Jagadish, H. V., and Faloutsos, C., "Efficient Retrieval of Similar Time Sequences Under Time Warping," In *Proc. of the 14th IEEE Int'l Conf. on Data Engineering*, Orlando, Florida, pp. 201-208, Feb. 1998.